Load Characterization and Low-order Approximation for Smart Metering Data in the Spectral Domain

***Abstract*— Smart metering data are providing new opportunities for various energy analyses at household level. However traditional load analyses based on time-series techniques are challenged due to the irregular patterns and large volume from smart metering data. This paper proposes a promising alternative to decompose smart metering data in the spectral domain, where i) the irregular load profiles can be characterized by the underlying spectral components, and ii) massive amount of load data can be represented by a small number of coefficients extracted from spectral components.**

This paper assesses the performances of load characterization at different aggregated levels by two spectral analysis techniques, using the discrete Fourier transform (DFT) and discrete wavelet transform (DWT). Results show that DWT significantly outperforms DFT for individual smart metering data while DFT could be effective at a highly aggregated level.

***Index Terms*—demand side response, discrete Fourier transform, discrete wavelet transform, load profiles, spectral analysis, smart grid, smart meter**

I. Introduction

Smart meters are the next generation of electricity meters and are rapidly developing across the world [1-3]. In the UK, the Department of Energy & Climate Change (DECC) aimed to install smart meters for all homes and small businesses by 2020 [4]. The transition will involve rolling out over 53 million smart meters. They can directly communicate the energy usage information of millions of individual customers to suppliers and network operators, thus improving the power system efficiency in the following forms: i) to support more efficient use of demand side response (DSR) [5]; ii) to inform the planning and operation of a smart grid [6, 7]; iii) to enhance the market settlement efficiency and accuracy [8].

Traditionally, customers' load information is characterized by *typical load profiles* (TLPs). Customers are classified based on similarities between load profiles in the time domain so that each group can be represented by a TLP. Different techniques such as clustering [9], classification [10] and neural networks [11] have been developed to classify customers. The essential condition for such techniques is the sufficient similarities in customer load profiles in the time domain. Therefore previous load characterization research usually focuses on industrial load and averaged/aggregated load, which are smooth, stable and share more similarities in the time domain.

However, smart metering data are naturally volatile and irregular, and thus are difficult to be characterized by TLPs. Figure 1 shows an example of the characteristics of smart metering data [1]. The data are from two real smart meters installed at two Irish residential households. They are depicted as the grey and black lines respectively, and figure 1 shows their daily load profiles over 10 days. Meanwhile, the traditional TLP used by the UK power industry is depicted by the single red line. As shown in the figure, traditional TLP cannot fully characterize the smart metering data in terms of the volatility and variances between days. As smart metering data inherit less similarities in the time domain, it is difficult to characterize them by time-series analysis [12].

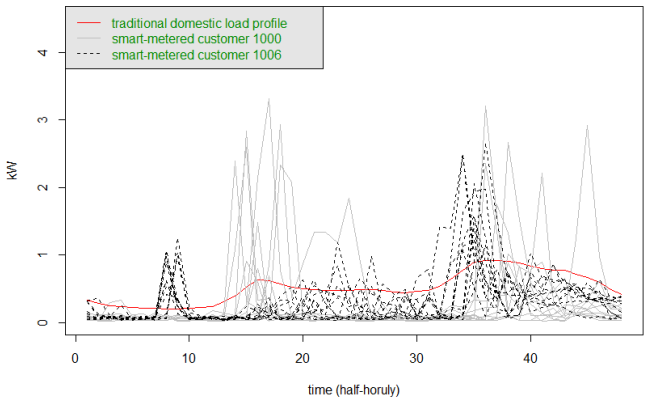

Figure 1. Comparison between traditional TLP and smart metered load profiles (data from Irish Smart Metering Project)

A promising alternative is spectral analysis, which has been applied in many different fields in power systems including power quality [13-15], power dispatch [16], forecasting in power systems [17], power system measurement [18], power system protection [19], power system transients [20], non-technical loss detection [21-23] and data compression [24].

The discrete Fourier transform (DFT) and the discrete wavelet transform (DWT) are two powerful and widely acknowledged techniques used in load characterization and low order approximation. The latter aims to use reduced number of spectral coefficients to approximate the original

load.

The major steps for characterizing load profiles by DFT are presented in [25]. The DC component of DFT can largely represent load factor, and selected harmonics can be used to describe the load shape [25-27]. The results show that average daily load profiles of customers can be adequately represented by a small set of frequency components. However, such a technique has only been applied on the system/aggregated levels. With the volatile and massive smart metering data, DFT is expected to suffer major limitations on load characterization.

DWT has mainly been studied for short-term load forecasting (STLF) at system level. Reference [28] emphasizes the advantage of the wavelet transform over DFT in that wavelets are able to capture short-duration pulses (e.g. particular events) and non-stationary features (e.g. seasonality within a year). Reference [29] adopts wavelets in the pre-processing stage to filter noise and redundant data. Reference [30] decomposes both load data and weather variables into low-frequency and high-frequency components, where low-frequencies can be precisely predicted. Reference [31] attempts to predict high-frequencies by a similar-day based neural network.

It can be seen most applications are limited to an aggregated level (i.e. grid level). The upcoming smart metering data will bring new opportunities to apply spectral analysis at a granular level. Applications have already been seen in data privacy [32], load pattern clustering [33] and non-intrusive load monitoring [34]. To our best knowledge, no research has been reported to assess performance of load characterization between DFT and DWT for smart metering data. There is limited effort in deploying the spectrum analysis for low-order approximation to substantially reduce the volume of smart metering data.

This paper proposes a novel method which simultaneously characterizes the irregular patterns and compresses the massive data from smart meters by low-order approximation. The DWT is adopted to effectively extract load characteristics in the spectral domain, at the same time substantially reducing the data size by using a limited number of components.

The main contributions of this paper are:

- development of a new low order approximation and load characterization method in the spectral domain for smart metering data;
- informing future smart metering research of the choice of spectral analysis techniques through two assessments: i) assessment of the applicability at the highly aggregated and disaggregated levels; ii) assessment of widely used techniques, DFT and DWT.

Key findings from this research are:

- the proposed DWT method decomposes smart metering data into more meaningful components and performs more effectively on low order approximation compared with DFT;
- however, the performance of DFT becomes more stable and superior to DWT when the load aggregation level is high.

The rest of the paper is organised as follows. Section II presents the spectral analysis techniques for load characterization. Section III proposes the assessment method for low-order approximation. Section IV briefly introduces the data used in this research. Assessment results for disaggregated load data are demonstrated in Section V and results for aggregated load are compared and discussed in Section VI. Conclusions are drawn in Section VII.

II. Spectral Analysis Techniques

The decomposition process can be treated as a transformation from one function into a different set of scaled basis functions. The basis functions of DFT are complex sinusoids of various frequencies while DWT adopts orthonormal wavelets [35, 36]. Reconstruction is basically the inverse transform; however, data can be compressed and characterized during this process.

Consider the daily load profile as a time series $\boldsymbol{s}=[s_0, s_1, \mathrm{K}\ s_{N-1}]$, where N is the daily sample size (N=144 for 10 minutes interval and N=48 for half-hourly interval). In order to compare the load decomposition on different aggregation levels, all daily load profiles are normalized to $\boldsymbol{b}=[b_0, b_1, \mathrm{K}\ b_{N-1}]$ according to its maximum daily load as shown in (1),

$$b_n=\frac{s_n}{\max_{\{i\ =\ 0,\ldots,\ N-1\}}\{\boldsymbol{s}_i\}},\quad n=0,\mathrm{K}\ N-1\ . \tag{1}$$

A. Discrete Fourier Transform

Using DFT, $\boldsymbol{b}$ can be transformed from the time domain to the frequency domain. The spectrum of $\boldsymbol{b}$ is shown by (2)

$$B_k=\sum_{n=0}^{N-1}b_n e^{-j\frac{2\pi kn}{N}},\quad k=0,\mathrm{K}\ N-1, \tag{2}$$

where $B_k=\beta_k e^{j\theta_k}$ is the frequency spectrum with magnitude of β_k and phase angle θ_k.

Using the inverse discrete Fourier transform (IDFT), the time series load profile $\boldsymbol{b}$ can be reconstructed by summing up the frequency components to $\boldsymbol{b}^r$:

$$b_n^r=\frac{1}{N}\sum_{k=0}^{N-1}B_k e^{j\frac{2\pi kn}{N}},\quad n=0,\mathrm{K}\ N-1\cdot \tag{3}$$

The complex coefficients can be merged in pairs forming cosine functions with different frequencies and initial phase angles (When N is an even number, the component of Nyquist frequency (k=N/2) is a triangular wave). For even N, the reconstruction of time series $\boldsymbol{b}$ can be expressed by (4):

$$b_n^r=\frac{\beta_0}{N}+\sum_{k=1}^{\frac{N}{2}-1}\frac{2\times\beta_k}{N}\cos\left(\frac{k2\pi n}{N}+\theta_k\right)+\frac{\beta_{\frac{N}{2}}}{N}\cos(\pi n+\theta_{\frac{N}{2}}),\ n=0,\mathrm{K}\ N-1. \tag{4}$$

B. Discrete Wavelet Transform

The Fourier transform is inefficient to decompose non-

stationary signals, whose frequency components vary over time and require a large number of harmonics to express volatile load profiles characterized as spikes or needle peaks.

Wavelet analysis mitigates the deficiency by introducing a wavelet that decays in a limited time window. It enables each component to have different scales and shifts over time. The decomposition process can be illustrated by Figure 2. The load profile is decomposed by high-pass and low-pass filters. The coefficients of the filters are determined by the choice of mother wavelet. The down-sampling process breaks down original load profiles into lower resolution components. A higher level of decomposition process will generate lower resolution components. The large-scale components are called "approximations" (A) while small-scale components are called "details" (D).

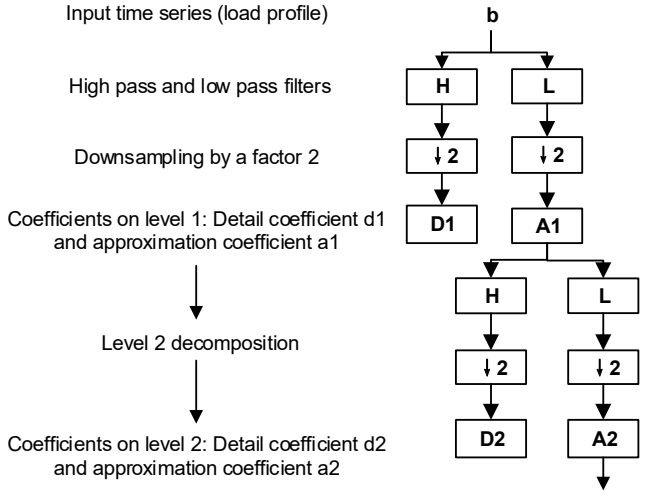

Figure. 2 Multi-resolution analysis by DWT

Through up-sampling and reconstruction filters, approximation component *A1* and detail component *D1* can be obtained. By this way, the original load profile can be represented by multi-resolution analysis (MRA) shown by (5):

$$\boldsymbol{b}^r = D1 + A1 = D1 + D2 + A2 = \mathrm{K} = \sum_{j=1}^{J} D_j + A_J \,, \quad (5)$$

where $\boldsymbol{b}^r$ is the reconstructed load profile; A_j and D_j are the approximation and detail components at level j, J is the total levels of decomposition.

Figure 3 gives an example of using DWT to decompose an individual customer's load profile. Plots from top to bottom are original load profile, D1, D2, D3 and A3 respectively. It is expressed in (6) and (7):

$$A_n^j = \sum_k a_{jk}\varphi_n^{jk} \;;\; j = 1,\ldots,J;\; n = 0,\mathrm{K}\; N-1 \quad (6)$$

$$D_n^j = \sum_k d_{jk}\phi_n^{jk} \;;\; j = 1,\ldots,J;\; n = 0,\mathrm{K}\; N-1 \quad (7)$$

Where A_n^j and D_n^j are the approximation component and the detailed component at level j and time index n. a_{jk} and d_{jk} are the scaling and mother wavelet coefficients corresponding to the scaling and mother wavelet functions φ_n^{jk} and ϕ_n^{jk} at level j and coefficient number k, expressed at time index n. For the real scaling functions, the scaling coefficients could be obtained by (8)

$$a_{jk} = \sum_{n=0}^{N-1} b_n \varphi_n^{jk} = 2^{-j/2} \sum_{n=0}^{N-1} b_n \varphi\left(\frac{n - k2^j}{2^j}\right) \quad (8)$$

Where time series $\boldsymbol{s} = [s_0, s_1, \mathrm{K}\; s_{N-1}]$ is a daily load profile and N is the daily sample size and $\varphi\left(\frac{n-k2^j}{2^j}\right)$ is the scaling function. Wavelet coefficients d_{jk} can be derived in a similar way but with mother wavelet ϕ_n^{jk}.

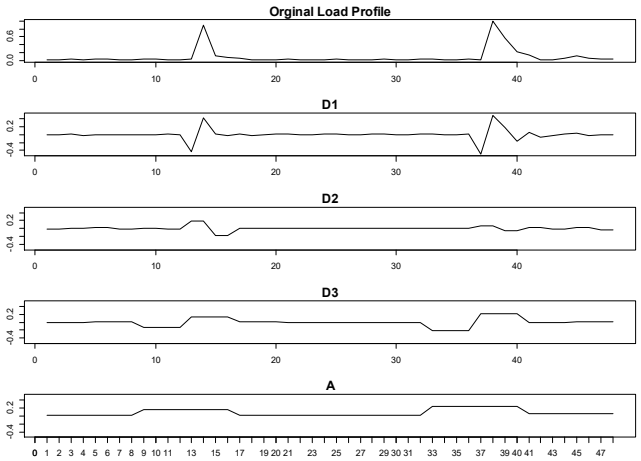

Figure. 3 Load profile decomposition by DWT (basis functions multiplied by coefficients)

This paper chooses *Haar* as the mother wavelet and a decomposition level of 3. We consider *Haar* is likely to be coherent with the nature of individual customer's load profile as the square wave can better portray the turn on-off of domestic appliances. In our case, the selection is denoted as the scaling function in (9) and mother wavelet in (10)

$$\varphi(t) = \begin{cases} 1 & 0 \le t < 1 \\ 0 & otherwise \end{cases} \quad (9)$$

$$\phi(t) = \begin{cases} 1 & 0 \le t < 1/2 \\ -1 & 1/2 \le t < 1 \\ 0 & otherwise \end{cases} \quad (10)$$

III. Assessment Method

DFT and DWT are compared as spectral representations of load profiles. The assessment is focused on feature representation in terms of: i) load characterization and ii) low-order approximation.

The assessment of DFT takes the following steps: i) data pre-processing: un-structured data sets are firstly cleaned, sense-checked, and organized into the same structure; daily

load profiles are normalized to a certain range; ii) decomposition: use DFT to decompose daily load profiles into frequency coefficients, including magnitudes and phase angles of all components; iii) load characterization: evaluate the coefficients in terms of composition, correlation and consistency (variations of the daily coefficients of the same customer over time); iv) low-order approximation: use a limited number of components, from one to all, to represent the original load profile.

The low-order approximation investigates the trade-off between profiling accuracy and data size reduction. It is noted that the zero-frequency component, which depicts the average loading level, usually dominates the magnitudes. Table I lists the DFT components of a sampled customer's load profile. As the frequency increases, the magnitudes of components dramatically drop. Aggregation of the first few DFT components is expected to capture the original load profile with high accuracy while the data size can be significantly reduced.

The idea is based on the assumption that the reconstruction is dominated by low-frequency (or large-scale) components. Using the first few coefficients will adequately resemble the original load profiles as they preserve the majority of the spectral energy, which is calculated as the sum squares of coefficients' magnitudes defined in (11) and (12):

$$E^{j}{}_{dft} = \begin{cases} \dfrac{\beta_0^2}{N} & j = 0 \\ \dfrac{\beta_0^2}{N} + \sum_{k=1}^{J} \dfrac{2\beta_k^2}{N} & 0 < j < \dfrac{N}{2} \\ \dfrac{\beta_0^2}{N} + \sum_{k=1}^{(N/2)-1} \dfrac{2\beta_k^2}{N} + \dfrac{\beta_{\frac{N}{2}}^2}{N} & j = \dfrac{N}{2} \end{cases} \quad (11)$$

$$E^{j}{}_{dwt} = \sum_{k} |a_{Jk}|^2 + \sum_{l=j}^{J} \sum_{k} |d_{lk}|^2 \qquad j = 1, \ldots, J \quad (12)$$

Where $E^{j}{}_{dft}$ is the accumulated DFT energy up to the jth harmonic. $E^{j}{}_{dwt}$ is the accumulated DWT energy up to the decomposition level of j.

Table I DFT coefficients of a sampled load profile

Relative Frequency	Amplitude	Phase
0 (DC)	0.72	0
1/48	0.173	1.82
2/48	0.151	1.55
3/48	0.027	-1.73
,…	……	……
23/48	0.003	0.24

The representativeness of reconstructed load profiles are evaluated by the following indices: Peak Magnitude Error Index (PMEI), Maximum Magnitude Error (MME), Mean Absolute Percentage Error (MAPE), Peak Time Error (PTE). All metrics are defined in (13)-(16), where $\boldsymbol{b}$ and $\boldsymbol{b}^r$ are the original and reconstructed load profiles; $t_{\max(\boldsymbol{b})}$ is the time when peak load occurs in the profile $\boldsymbol{b}$. This paper follows the same criteria for reconstruction assessment as in [25]. A reconstructed load profile is considered satisfactory if the PMEI, MME and MAPE are all below 5% and PTE is shorter than 2 hours.

$$PMIE = \left| \frac{\max_i(b_i) - \max_i(b_i^r)}{\max_i(b_i)} \right| \times 100, \quad (13)$$

$$MME = \max_i \left(\left| \frac{b_i - b_i^r}{b_i} \right| \right) \times 100, \quad (14)$$

$$MAPE = \sum_{i=1}^{N} \left(\left| \frac{b_i - b_i^r}{b_i} \right| \right) \times \frac{100}{N}, \quad (15)$$

$$PTE = t_{\max(\boldsymbol{b})} - t_{\max(\boldsymbol{b}^r)} \quad (16)$$

The assessment of DWT follows similar steps to those of DFT. However, the low order approximation method for DWT is modified. Besides the use of a limited number of components, it is noted that DWT components, especially the small-scale ones, have very low magnitudes through most of the time windows. Thus, the additional method for DWT low order approximation is to remove the low-demand periods of each scale. Coefficients below threshold will be set as zeros. By this way, the number of non-zero coefficients can be significantly reduced.

IV. Data Description

The evaluation is implemented at different aggregated levels including for individual customers, averaged customers and low voltage (LV) substations. Two sets of data respectively from smart grid and smart meter projects are assessed in this paper. The smart grid demonstration project, Low Voltage Network Templates Project [37] is jointly commissioned by Western Power Distribution (WPD) in the UK and the UK's regulator - the Office of Gas and Electricity Markets (Ofgem). WPD deployed monitoring equipment at 800 HV/ LV substations and over 3,500 ends of LV feeders collecting network performance data. The variable data collection is on a 10-minute interval over the course of one year (2012-2013), including three-phase voltage, current and real power delivered at HV/LV substations.

The smart metering data are from the Irish smart meter trial project [1]. There are 6369 customers with half-hourly demand recorded over one and a half years (2009-2011). For LV substations, daily load profiles will be assessed. For individual customers, both monthly average and daily load profiles will be assessed. The data from LV substations are relatively smoother than smart metering data, which are extremely volatile.

V. Results For Disaggregated Load

A. Individual Customer

The most unique characteristic of daily load profiles of

individual customers is volatility. Figure 4 shows the daily load profiles of customer 1002 in July 2012. The significant volatility of daily load profiles (grey) makes it inaccurate to represent them by average (red). It is also difficult to use any random day to represent the month unless some meaningful information can be extracted from these irregular load profiles.

B. Load Characterization

Theoretically, it is suggested that the volatile load profiles can be decomposed into more stable and meaningful components by DWT compared with DFT. The reason is that the DFT basis functions are periodic and stationary. It may require many high-frequency components to resemble the volatility of original load profiles. The shift of "needle peaks" from original load profiles may result in large variation in the DFT coefficients. On the other hand, DWT is dynamic on both frequency and time domain, which enables it to capture the sudden spikes and hold the underlying trend at the same time.

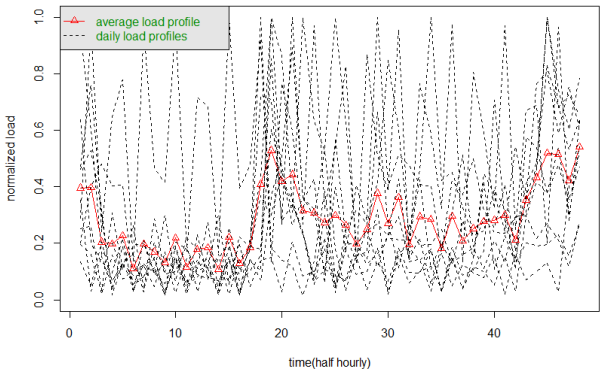

Figure. 4 Daily load profiles of customer 1002 in July 2012

In this assessment, DFT and DWT are both used to decompose the daily load profiles of 6369 customers through a year. Figure 5 illustrates the decomposition by DFT. The volatile black line is the real load profile of a sample customer. The red line is its DC component representing the first part in (4). The remaining colourful lines are the AC components with different frequencies. Summing up these components can get artificial time series that resemble the original load profile.

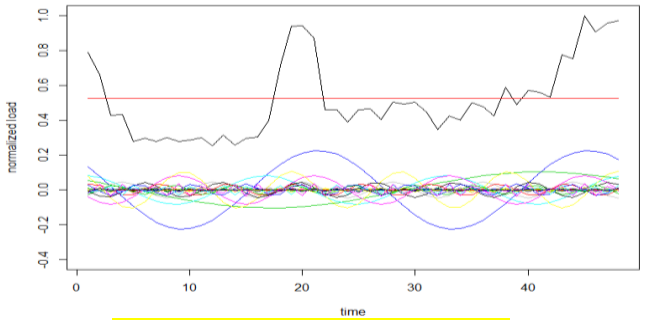

Figure. 5 Load profile decomposition by DFT

Figure 6 shows the decomposition scales from DWT. Each daily load profile in Figure 5 is decomposed by DWT into 4 components: A3, D3, D2, D1, with scale from large to small. The observations are as follows.

a) The approximation (A3) components describe the underlying trend of daily load profiles. For the same customer, A3 components are generally consistent through different days. With further classification of seasons, months and day types, the similarity is expected to increase. In Figure 6, Customer 1002 shows a fundamental usage pattern of "double-peak" in July.

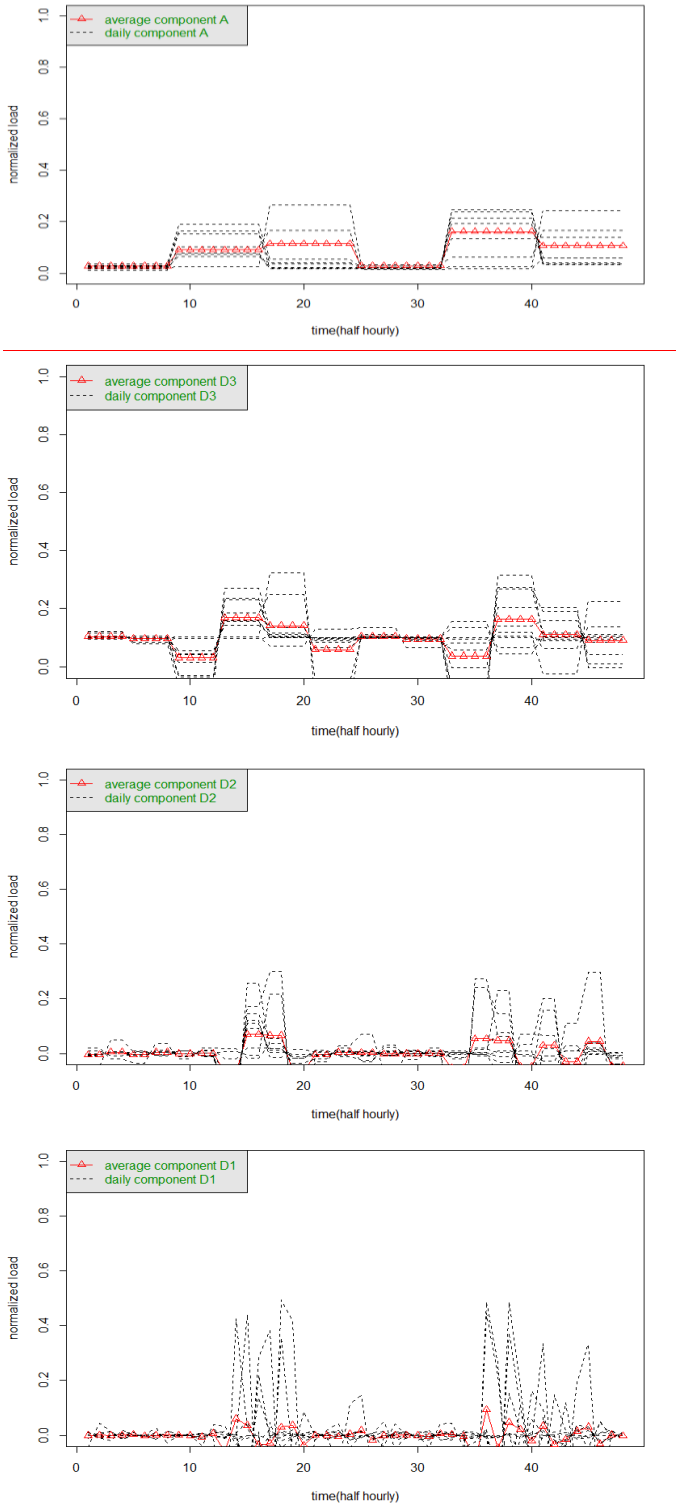

Figure. 6 Decomposition scales from DWT for customer 1002 (coefficients multiplied by basis wavelets)

b) D3 and D2 components represent more random activities and short-interval usage (e.g. kettles). Figure 6 clearly sees the low-demand time from 1 a.m. to 8 a.m. (sleeping time) and busy time in the morning and evening.

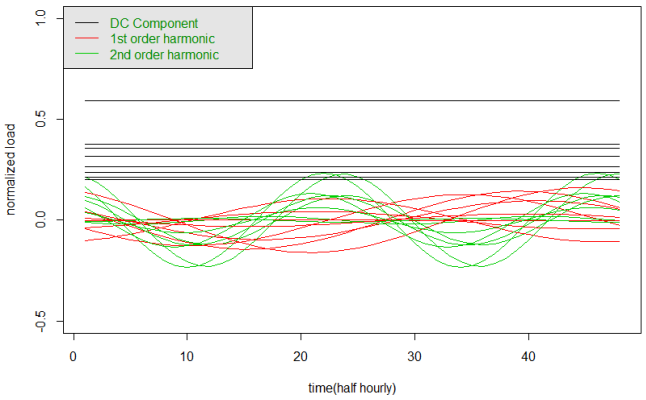

Figure. 7 Periodical sinusoidal components from DFT(basis sinusoids multiplied by coefficients)

c) The D1 scale is the smallest scale. It contains random spikes which are possibly caused by the turn-on of some appliances. It is also noted that some of the D1 components are quite periodical, likely to represent white goods such as refrigerators.

By contrast, the periodical sinusoidal components from DFT reveal less information as shown in Figure 7.

It is found that DWT is better at load characterization by two advantages: i) DWT decomposes the load to more meaningful components; ii) the DWT coefficients are more consistent within the same customer through days.

C. *Low-order approximation*

A reduced number of the transformed coefficients can be used to re-construct the original load profile with errors. Hence, another assessment is to evaluate the trade-off between representativeness of the reconstructed load profile and low order approximation.

The assessment has been conducted on all load profiles, reconstructing from low to high frequencies (large to small scales) by both DFT and DWT. Figure 8 shows the accumulated spectral energy by keeping different numbers of coefficients from DFT and DWT. We calculated the energy in the time domain, which gives the same results as (11) and (12). As shown in the figure, keeping all 48 coefficients, both methods will preserve 100% of the spectral energy in original load profiles while the first coefficient alone contains 20% of the original spectral energy. The observations are as follows.

a) The first coefficient of both methods has around 24% spectral energy of the original load profile, which is consistently close with the load factor (average/peak) of the original load profile. It is expected because the DC component signifies the mean value of the original signal, and in our case (normalized load profile with peak "1") the load factor is exactly the mean.

b) The large-scale component of DWT contains more spectral energy, with over 99% spectral energy after first 6 coefficients. The spectral energy is spread more evenly over DFT coefficients, reaching only 90% after 24 coefficients. It shows that with the same low order approximation, the DWT reconstruction will preserve more spectral energy of the original load profile.

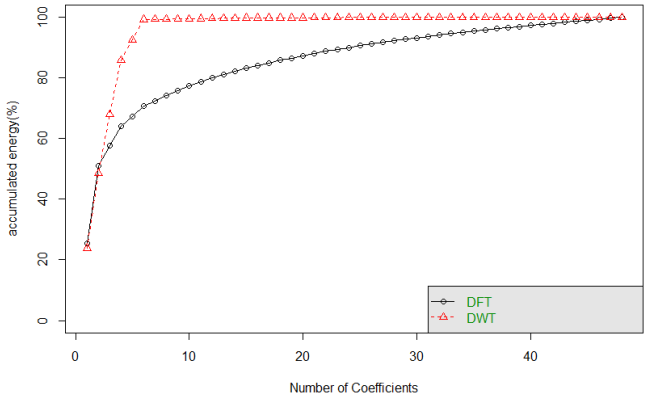

Figure. 8 Accumulated spectral energy by keeping different numbers of coefficients

The data size of DWT can be further reduced by eliminating all "near-zero" coefficients within a scale. Especially the small-scale components of DWT, which are likely to see low-demand for long time and only several spikes over a day, contains many coefficients close to zero. Eliminating those coefficients will hardly affect the reconstruction accuracy meanwhile reducing the data size considerably.

To further compare the low order approximation ability of DFT and DWT, an extensive comparison is conducted between original load profiles and reconstructed load profiles. Four indices (PMEI, MME, MAPE and PTE) widely used in load profiling are adopted here as introduced in (13)-(16).

6369 customers' daily load profiles are reconstructed with different sizes of reduced data. The test is to find the minimum data size required to meet the reconstruction accuracy. In other words, the aim is to find the possible lowest order approximation while keeping the reconstruction error under the threshold. In this paper, as noted above the error threshold is set to be 5% for PMEI, MME, MAPE and 2 hours for PTE. It follows the previous studies [6] so that the results are comparable.

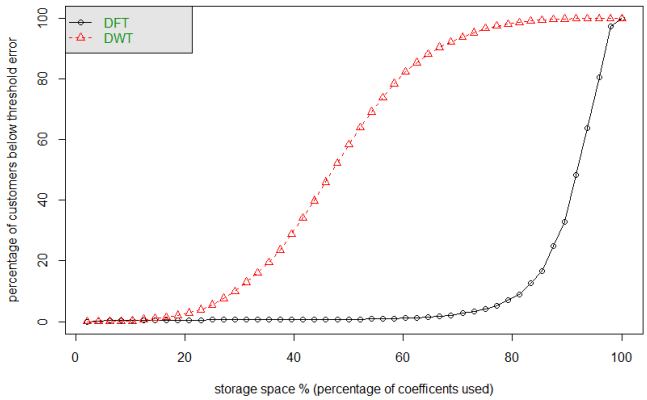

Figure. 9 Percentage of customers who can be reconstructed under the threshold error with different data size

Figure 9 shows the percentage of customers who can be reconstructed under the threshold error with different data size. The abscissa axis is the data size (100%=48 coefficients) used to reconstruct the load profiles. For example, using half of the DFT coefficients, only 0.8% of the total customers' load profiles (about 46 customers) can be reconstructed with an error below threshold. However, using half of the DWT coefficients, 58% of the customers' load profiles can be satisfactorily reconstructed. Other main findings are as follows.

a) The reconstruction can hardly meet the accuracy requirements with less than 20% of the coefficients for both techniques. The pass rate starts to increase when using more than 20% of DWT data. However, the DFT pass rate remains low until using more than 80% of its coefficients. For volatile load profiles, DFT needs relatively complete high-frequency component sets to resemble the sudden spikes while DWT can handle that with only a few small-scale coefficients.

b) Even with all of the DFT coefficients below Nyquist frequency (47/48), still 2.8% (174 out of 6369) of the customers' load profiles cannot be reconstructed below the threshold error. However, with 47 of the DWT coefficients, all load profiles can be successfully recovered.

c) The largest gap between the 2 techniques occurs at 75% of the data size. Using 75% of DWT coefficients can recover 96.7% of the original load profiles. However, only 4.2% of the original load profiles are recovered by 75% of the DFT coefficients. The difference is as high as 92.5%. The fundamental reason is probably that the natural shapes of smart metering load profiles are more coherent with the *Haar* wavelet than with sinusoidal waves.

VI. Results For Aggregated Load

Different applications of load profiles focus on different aggregation levels. Some tariff design is based on aggregation over time while network planning pays more attention to aggregation over customers. We roll out similar assessments as in VI but on different aggregation levels. For aggregation over time, monthly average load profiles of 6369 smart metering customers are tested. For aggregation over customers, the daily load profiles from 800 LV substations are used.

A. Monthly Averaged Load Profiles

Figures 10 and 11 show the difference between monthly average and daily individual load profiles with heavily reduced data (80% reduction). As shown in the figures, the performance of DFT is significantly compromised in low-order approximation while the performance of DWT is much less affected. The black line in Figure 10 is the daily load profile of customer ID 1000 on 1st July 2012. The red line is the reconstructed load profile by the DC and first two harmonic components (5 coefficients). The blue line is the reconstructed load profile by the largest 6 DWT coefficients. Clearly, with similar data size, reconstruction of DWT is much better than that of DFT.

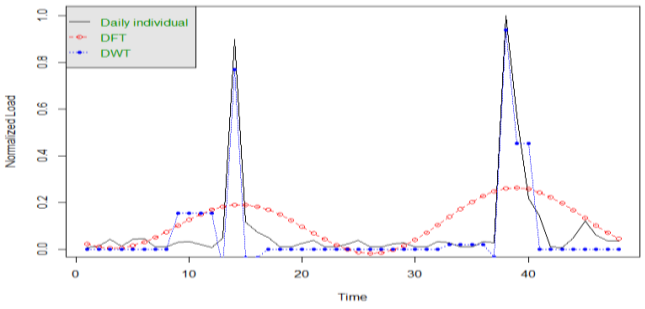

Figure. 10 Daily individual load profile and low order reconstructions by DFT and DWT

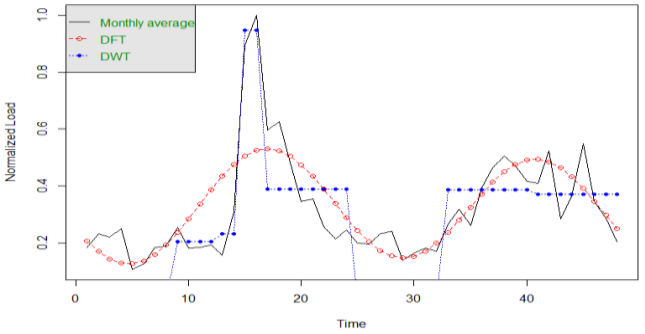

Figure. 11 Monthly average load profile and low order reconstructions by DFT and DWT

In Figure 11, the black line is the average load profile of customer ID 1000 in July. It is smoother than the daily load profile. Using the same reduced data size to reconstruct the average load profile, DFT shows a much better performance compared with that on daily load. Although DWT still resembles the original load profiles better than DFT, the gap is substantially narrowed. This is also illustrated by Figure 12, which is a comparable plot to Figure 9. It is the successful reconstruction rate for monthly average load profiles with different data sizes. The performance of DWT is very similar with that of daily load profiles. However, DFT shows an overall improvement. Using 80% of the DFT coefficients can recover 5.7% of the daily load profiles, but 48.5% of the monthly average load profiles.

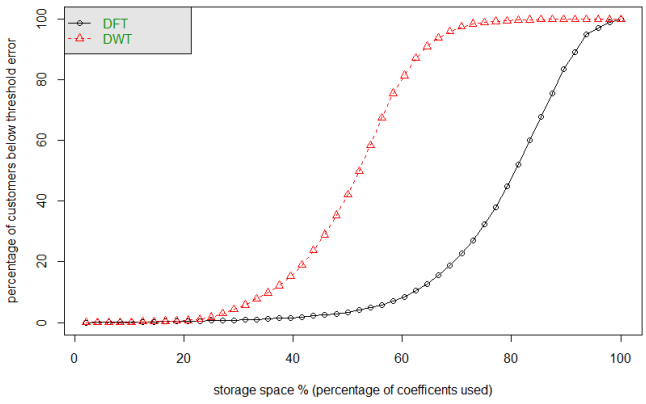

Figure. 12 Percentage of customers who can be reconstructed under the threshold error with different data size (monthly average load profiles)

B. LV Substation Load Profiles

The daily load profiles at LV substations are representatives

for aggregated load over customers. The assessment shows that when the load profiles are granular, DWT constantly performs better at low order approximation; however, DFT improves significantly as the aggregation level increases.

Substations are assessed by their customer sizes. In order to demonstrate a continuous change of customer size, some individual customers load profiles are added onto the substation artificially. Figure 13 shows the average minimum data required to reconstruct load profiles of different customer groups. As the customer size increases, load profiles are more aggregated and smooth. Naturally, the data size required to reconstruct the load profiles decrease for both techniques. The interesting findings are as follows.

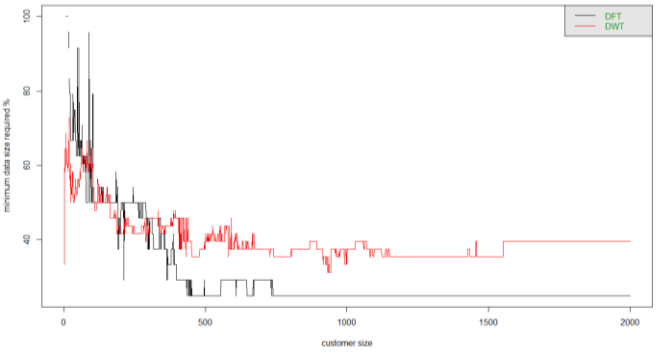

Figure. 13 Average minimum data required to reconstruct load profiles from DFT and DWT coefficients for different customer groups (PMEI, MME, MAPE< 5% and PTE<2 hours)

a) When the customer size is small, DWT is generally superior to DFT. However, when the customer size increases over 400, DFT requires less coefficients than DWT in terms of reconstruction.

b) Further, when the customer size is larger than 450, the DFT steadily requires only the DC and the first harmonic components (first 3 components, 12% of the coefficients) to fulfil the reconstruction. This figure further becomes constant when over 700 customers.

c) In contrast, DWT on average requires 48% of the data to reconstruct small group of customers' load profiles. For larger groups, it on average requires 17 (35%) of its coefficients with some fluctuations.

VII. Conclusion

This paper presents a load characterization method for smart metering data based on spectral analysis. Assessment of the low order approximation performance has been conducted for load profiles at aggregated and disaggregated levels. The two tested spectral analysis techniques, DFT and DWT can be viewed as two extreme cases. In DWT, *Haar* is chosen as a compact wavelet while the sinusoidal wave in DFT gives a global support. The key findings are as follows.

i) At disaggregated level, DWT can characterize load profiles into more meaningful and consistent components compared with DFT.

ii) At disaggregated level, DWT is also more effective in terms of low order approximation than DFT. DWT can reconstruct the original daily load profiles using less coefficients while maintaining high representativeness.

iii) However, at more aggregated level, the performance of DFT is substantially improved. The performance of DFT becomes stable and superior to DWT when the aggregation level is sufficiently high.

Based on the results, DFT could be effective for load profiling at high-aggregated level while DWT is more promising at granular level. The results of this paper will provide a valuable reference on the choice of techniques at different aggregation levels. The case study of load characterization in this paper could support the following research:

- Forecasting: the load at household level is extremely volatile and thus difficult to be forecasted. Our work could be used to filter out those volatilities and focus on the residues.
- DSR: the design of demand side response at domestic homes is usually based on TLPs. The error in TLPs may causes inefficient operation and control for domestic household. Based on the results, future work will focus on the classification and load profiling of the diverse end customers.

References

[1] C. f. E. R. (CER). CER Smart Metering Project [Online]. Available: http://www.ucd.ie/issda/data/commissionforenergyregulationcer/

[2] "Managing big data for smart grids and smart meters," USAMay 2012.

[3] "Utility-scale Smart Meter Deployments," The Edison Foundation Institue for Electric InnovationSpe 2014.

[4] "Smart Metering Implementation Programme," Department of Energy & Climate Change, London, UKDec 2014.

[5] P. Siano and D. Sarno, "Assessing the benefits of residential demand response in a real time distribution energy market," *Applied Energy,* vol. 161, pp. 533-551, 1/1/ 2016.

[6] Q. Sun, H. Li, Z. Ma, C. Wang, J. Campillo, Q. Zhang, *et al.*, "A Comprehensive Review of Smart Energy Meters in Intelligent Energy Networks," *IEEE Internet of Things Journal,* vol. 3, pp. 464-479, 2016.

[7] H. Arasteh, M. S. Sepasian, and V. Vahidinasab, "An aggregated model for coordinated planning and reconfiguration of electric distribution networks," *Energy,* vol. 94, pp. 786-798, 1/1/ 2016.

[8] S. S. S. R. Depuru, L. Wang, and V. Devabhaktuni, "Smart meters for power grid: Challenges, issues, advantages and status," *Renewable and Sustainable Energy Reviews,* vol. 15, pp. 2736-2742, 8// 2011.

[9] G. Chicco, R. Napoli, and F. Piglione, "Comparisons among clustering techniques for electricity customer classification," *Power Systems, IEEE Transactions on,* vol. 21, pp. 933-940, 2006.

[10] Z. Tiefeng, Z. Guangquan, L. Jie, F. Xiaopu, and Y. Wanchun, "A New Index and Classification Approach for Load Pattern Analysis of Large Electricity Customers," *Power Systems, IEEE Transactions on,* vol. 27, pp. 153-160, 2012.

[11] D. Gerbec, S. Gasperic, I. Smon, and F. Gubina, "Determining the load profiles of consumers based on fuzzy logic and probability neural networks," *Generation, Transmission and Distribution, IEE Proceedings-,* vol. 151, pp. 395-400, 2004.

[12] Z. Chen, S. Heng, L. Ran, and L. Furong, "Demand side response performance assessment: An impact analysis of load profile accuracy on DSR performances," in *2015 IEEE Power & Energy Society General Meeting*, 2015, pp. 1-5.

[13] F. B. Costa, "Boundary Wavelet Coefficients for Real-Time Detection of Transients Induced by Faults and Power-Quality Disturbances," *IEEE Transactions on Power Delivery,* vol. 29, pp. 2674-2687, 2014.

[14] F. A. S. Borges, R. A. S. Fernandes, I. N. Silva, and C. B. S. Silva, "Feature Extraction and Power Quality Disturbances Classification Using Smart Meters Signals," *IEEE Transactions on Industrial Informatics,* vol. 12, pp. 824-833, 2016.

[15] S. Ali, K. Wu, K. Weston, and D. Marinakis, "A Machine Learning Approach to Meter Placement for Power Quality Estimation in Smart Grid," *IEEE Transactions on Smart Grid,* vol. 7, pp. 1552-1561, 2016.

[16] H. Chao-Ming and H. Yann-Chang, "Combined wavelet-based networks and game-theoretical decision approach for real-time power dispatch," *IEEE Transactions on Power Systems,* vol. 17, pp. 633-639, 2002.

[17] A. S. Pandey, D. Singh, and S. K. Sinha, "Intelligent Hybrid Wavelet Models for Short-Term Load Forecasting," *IEEE Transactions on Power Systems,* vol. 25, pp. 1266-1273, 2010.

[18] S. Avdakovi, x, E. Be, x, irovi, x, *et al.*, "Generator Coherency Using the Wavelet Phase Difference Approach," *IEEE Transactions on Power Systems,* vol. 29, pp. 271-278, 2014.

[19] E. Y. Hamid and Z. I. Kawasaki, "Wavelet-based data compression of power system disturbances using the minimum description length criterion," *IEEE Transactions on Power Delivery,* vol. 17, pp. 460-466, 2002.

[20] W. A. Wilkinson and M. D. Cox, "Discrete wavelet analysis of power system transients," *IEEE Transactions on Power Systems,* vol. 11, pp. 2038-2044, 1996.

[21] S. C. Huang, Y. L. Lo, and C. N. Lu, "Non-Technical Loss Detection Using State Estimation and Analysis of Variance," *IEEE Transactions on Power Systems,* vol. 28, pp. 2959-2966, 2013.

[22] A. Jindal, A. Dua, K. Kaur, M. Singh, N. Kumar, and S. Mishra, "Decision Tree and SVM-Based Data Analytics for Theft Detection in Smart Grid," *IEEE Transactions on Industrial Informatics,* vol. 12, pp. 1005-1016, 2016.

[23] P. Jokar, N. Arianpoo, and V. C. M. Leung, "Electricity Theft Detection in AMI Using Customers’ Consumption Patterns," *IEEE Transactions on Smart Grid,* vol. 7, pp. 216-226, 2016.

[24] J. Ning, J. Wang, W. Gao, and C. Liu, "A Wavelet-Based Data Compression Technique for Smart Grid," *IEEE Transactions on Smart Grid,* vol. 2, pp. 212-218, 2011.

[25] Z. Shiyin and K. Tam, "A Frequency Domain Approach to Characterize and Analyze Load Profiles," *Power Systems, IEEE Transactions on,* vol. 27, pp. 857-865, 2012.

[26] K. Jungsuk, J. Flora, and R. Rajagopal, "Household Energy Consumption Segmentation Using Hourly Data," *Smart Grid, IEEE Transactions on,* vol. 5, pp. 420-430, 2014.

[27] M. P. Tcheou, L. Lovisolo, M. V. Ribeiro, E. A. B. da Silva, M. A. M. Rodrigues, J. M. T. Romano, *et al.*, "The Compression of Electric Signal Waveforms for Smart Grids: State of the Art and Future Trends," *Smart Grid, IEEE Transactions on,* vol. 5, pp. 291-302, 2014.

[28] A. J. R. Reis and A. P. Alves da Silva, "Feature extraction via multiresolution analysis for short-term load forecasting," *Power Systems, IEEE Transactions on,* vol. 20, pp. 189-198, 2005.

[29] Z. A. Bashir and M. E. El-Hawary, "Applying Wavelets to Short-Term Load Forecasting Using PSO-Based Neural Networks," *Power Systems, IEEE Transactions on,* vol. 24, pp. 20-27, 2009.

[30] A. S. Pandey, D. Singh, and S. K. Sinha, "Intelligent Hybrid Wavelet Models for Short-Term Load Forecasting," *Power Systems, IEEE Transactions on,* vol. 25, pp. 1266-1273, 2010.

[31] C. Ying, P. B. Luh, G. Che, Z. Yige, L. D. Michel, M. A. Coolbeth, *et al.*, "Short-Term Load Forecasting: Similar Day-Based Wavelet Neural Networks," *Power Systems, IEEE Transactions on,* vol. 25, pp. 322-330, 2010.

[32] D. Engel, "Wavelet-based load profile representation for smart meter privacy," in *Innovative Smart Grid Technologies (ISGT), 2013 IEEE PES*, 2013, pp. 1-6.

[33] K. Mets, F. Depuydt, and C. Develder, "Two-Stage Load Pattern Clustering Using Fast Wavelet Transformation," *IEEE Transactions on Smart Grid,* vol. PP, pp. 1-1, 2015.

[34] Y. C. Su, K. L. Lian, and H. H. Chang, "Feature Selection of Non-intrusive Load Monitoring System Using STFT and Wavelet Transform," in *e-Business Engineering (ICEBE), 2011 IEEE 8th International Conference on*, 2011, pp. 293-298.

[35] A. M. Davood Rafiei "Efficient Retrieval of Similar Time Sequences Using DFT," in *Proc. Int'l Conf. Foundations of Data Organizations and Algorithms*.

[36] A. M. M. Kociołek, M. Strzelecki P. Szczypiński "Discrete wavelet transform – derived features for digital image texture analysis," in *Proc. of Interational Conference on Signals and Electronic Systems*, Lodz, Poland, 2001, pp. 163-168.

[37] (2012). *LV Network Templates*. Available: http://www.ofgem.gov.uk/Networks/ElecDist/lcnf/stlcnp/year1/lv-network-templates/Pages/index.aspx